

Torsional Network Model: Normal Modes in Torsion Angle Space Better Correlate with Conformation Changes in Proteins

Raul Mendez and Ugo Bastolla*

Centro de Biología Molecular “Severo Ochoa,” (CSIC-UAM), Cantoblanco, 28049 Madrid, Spain

(Received 23 November 2009; published 3 June 2010)

We introduce the torsional network model (TNM), an elastic network model whose degrees of freedom are the torsion angles of the protein backbone. Normal modes of the TNM displace backbone atoms including C_β maintaining their covalent geometry. For many proteins, low frequency TNM modes are localized in torsion space yet collective in Cartesian space, reminiscent of hinge motions. A smaller number of TNM modes than anisotropic network model modes are enough to represent experimentally observed conformation changes. We observed significant correlation between the contribution of each normal mode to equilibrium fluctuations and to conformation changes, and defined the excess correlation with respect to a simple neutral model. The stronger this excess correlation, the lower the predicted free energy barrier of the conformation change and the fewer modes contribute to the change.

DOI: 10.1103/PhysRevLett.104.228103

PACS numbers: 87.14.et

Protein flexibility is essential for function, allowing structural rearrangements to propagate in allosteric regulation [1], and it is believed to play an important role in protein structure evolution [2]. A proper description of flexibility would be highly desirable both for improving computational drug design through flexible protein-ligand docking [3] and for refining homology models of protein structures [4]. Nevertheless, this is a very challenging task because of the very long time scales involved in functional motions and the lack of a computationally manageable, yet detailed enough statistical mechanical model of proteins. In this context, normal mode analysis [5] (NMA) provides a simple analytical description of the thermal dynamics of proteins in their native state. Normal modes calculations require an effective free energy function and a native structure that is placed at a local minimum of this energy function. However, energy minimization through the effective energy functions adopted in molecular dynamics can drive the native structure several Å away from the experimental one. An alternative approach, proposed by Tirion [6] and followed by several authors [7–10], consists in adopting as a starting point the experimentally known native structure and designing an effective free energy function that places such a structure at the absolute minimum. Such an approach, called the elastic network model (ENM), is analogous to GO models, which describe the statistical mechanics of protein folding using as only input the experimentally known native structure. Despite their simplicity, GO models are sometimes superior to more realistic but more complicated models in predicting protein folding kinetics [11]. Similarly, low frequency ENM normal modes, describing collective fluctuations in the native state, have highly significant overlap with experimentally observed conformation changes [12–15] and with evolutionary deformations of protein structures [16]. The success of ENMs indicates that a lot of information about

protein flexibility is contained in the topology of the native state.

In one of the earliest applications of NMA to proteins, it was proposed that NMA should consider only torsional degrees of freedom, which do not modify the covalent geometry [17]. However, this approach has not been pursued in ENMs, apart from a recent work in which the group of Ma developed an ENM based on pseudodihedral angles [18]. Here we propose the torsional network model (TNM), a new ENM that uses the torsion angles of the protein backbone as degrees of freedom, combining the topology of the native structure with the constraints imposed by the covalent geometry of proteins. We use the same effective potential as in previous ENMs. This potential is minimally frustrated [19], so that all native interactions are at a local minimum, and all the force constants are equal. We denote by $\{\vec{r}_i\}$, $i = 1 \cdots n$ the coordinates of n representative atoms having masses m_i , by $\{\vec{r}_i^0\}$ the corresponding equilibrium positions, and set $\vec{r}_{ij} = |\vec{r}_i - \vec{r}_j|$. With this notations, the effective potential free energy is

$$V \approx \frac{\gamma}{2} \sum_{ij} C_{ij} (|\vec{r}_{ij}| - |\vec{r}_{ij}^0|)^2, \quad (1)$$

and the kinetic energy is $T^{(r)} = \frac{1}{2} \sum_i m_i |\dot{\vec{r}}_i|^2$. Native interactions between atoms i and j are represented through a binary contact matrix C_{ij} equal to one if the interacting atoms are below a threshold distance in the native structure, zero otherwise. We examined thresholds ranging from 4 to 20 Å. Real-valued $C_{ij} \propto (r_{ij}^0)^{-e}$ may be considered as well. We analyzed 3 kinds of contacts: (1) based on C_α atoms, as in most previous ENMs; (2) based on C_β atoms (C_α for glycine); (3) based on all atoms, in which case for each residue pair only the two closest heavy atoms interact. Representative atoms may be C_α , C_β or all backbone atoms (including C_β and O).

In our model we consider only torsional degrees of freedom, forbidding modifications of bond angles and bond lengths. Torsional displacements affect the potential energy exactly as in the corresponding anisotropic network model (ANM). We denote θ_a the standard torsion angle φ_k if $a = 2k - 1$ and ψ_k if $a = 2k - 2$, \vec{e}_a the corresponding axis, and Ω_a the set of atoms upstream of \vec{e}_a , being Ω the set of all representative atoms. By convention, rotations around the axis \vec{e}_a only affect atoms in Ω_a . The derivative of the Cartesian vectors is $\partial \vec{r}_i / \partial \theta_a = \chi_{ia} \vec{e}_a \times (\vec{r}_i - \vec{s}_a)$, where χ_{ia} is 1 if $i \in \Omega_a$, zero otherwise, \times denotes vector product, and \vec{s}_a is the origin of axis \vec{e}_a . The Jacobian matrix J restricted to internal degrees of freedom is calculated imposing the Eckart conditions [20,21] as

$$\vec{J}_{ia} \equiv \frac{\partial \vec{r}_{i,\text{int}}}{\partial \theta_a} = \vec{e}_a \times (\vec{r}_i - \vec{s}_a) \chi_{ia} + \vec{A}_a \times \vec{r}_i + \vec{t}_a, \quad (2)$$

and the Hessian of the potential and kinetic energy are given by

$$U_{ab}^{(\theta)} \equiv \frac{\partial^2 V}{\partial \theta_a \partial \theta_b} = (J^t U^{(r)} J)_{ab}, \quad (3)$$

$$T_{ab}^{(\theta)} \equiv \frac{\partial^2 T^{(r)}}{\partial \theta_a \partial \theta_b} = \sum_i m_i \vec{J}_{ia} \cdot \vec{J}_{ib} \equiv (J^t M J)_{ab}, \quad (4)$$

where superscript t indicates the transpose matrix, M is the diagonal mass matrix, and $U^{(r)}$ is the Cartesian Hessian matrix of the potential energy. The normal modes of the TNM correspond to synchronized rotations with angular amplitude v_a^α around axis a , and solve the generalized eigenvalue equation $U^{(\theta)} v^\alpha = \omega_\alpha^2 T^{(\theta)} v^\alpha$. The corresponding Cartesian modes are obtained at first order in v_a^α as

$$\vec{x}_i^\alpha = \sum_a \vec{J}_{ia} v_a^\alpha. \quad (5)$$

Cartesian modes are orthogonal with respect to the norm defined by the kinetic energy matrix, i.e., $\sum_i m_i \vec{x}_i^\alpha \cdot \vec{x}_i^\beta = \delta_{\alpha\beta}$. The contribution of mode α to thermal fluctuations is $z^\alpha \propto \sum_i m_i \langle \delta r_i^2 \rangle_\alpha = (k_B T / \omega_\alpha^2) \sum_i m_i |\vec{x}_i^\alpha|^2 \propto \omega_\alpha^{-2}$.

We compare our model with the ANM [9], whose degrees of freedom are the Cartesian coordinates of the α carbons. An unwanted feature of the ANM is that modifying covalent bond lengths and bond angles has the same energetic cost as modifying soft degrees of freedom such as torsion angles, so that ANM excitations do not conserve the correct covalent geometry and secondary structure; i.e., they are not proteinlike. This problem has been addressed by strengthening the force constant of covalent bonds [8,22]. Ma and co-workers [18] recently introduced a model whose degrees of freedom are pseudo-torsion angles, which do not modify the distance between consecutive C_α . However, torsion angles offer the advantage that they allow us to represent all backbone atoms including C_β . Although torsion angles were adopted in one of the earliest studies of protein normal modes [17] and

recently reconsidered for modeling elastic deformations of proteins [23], they have not been applied in the context of ENM. TNM modes have three clear advantages with respect to ANM ones. (1) The displacements of atoms other than α carbons are accurately predicted without additional cost. (2) The number of degrees of freedom, hence of normal modes, is smaller: $2L - 2$ instead of $3L - 6$, where L is the number of residues. The computation is therefore accelerated by a factor $(3/2)^3 \approx 3.4$. (3) The structures generated applying small amplitude perturbations conserve the correct covalent geometry up to first order in $\Delta \theta_a$, which can be useful for the goal of building proteinlike structures [22]. Besides these obvious points, we compared TNM and ANM modes concerning their collectivity and their correlation with conformation changes.

Following Ref. [24], we measure the collectivity of normal mode α as the exponential of the Shannon entropy of its squared components, $\kappa[\{p_k^\alpha\}] = \exp(-\sum_k p_k^\alpha \log(p_k^\alpha))$. κ can be roughly interpreted as the number of objects with significant weight p_k^α . For the torsional collectivity we use $p_k^\alpha = (v_k^\alpha)^2 / \sum_j (v_j^\alpha)^2$ and for the Cartesian collectivity we use $p_k^\alpha = m_k |\vec{x}_k^\alpha|^2$. Torsional and Cartesian collectivities are correlated for high frequency modes. Nevertheless, for several proteins we observed that the lowest frequency TNM modes are collective in Cartesian space; i.e., a significant fraction of atoms are significantly displaced, but they are localized in torsional space; i.e., only a few angles are significantly modified. This clearly nonrandom pattern may help to select hinge modes [25]. We plot in Fig. 1 the mean torsional versus Cartesian collectivity of the four lowest frequency modes both for the TNM and the ANM. The torsional deviations of ANM modes and experimental conformation changes cannot be calculated simply by subtracting the corresponding angles, since the rotation axes are different, and the errors introduced neglecting this fact amplify when they propagate along the chain. We estimate them as the angles whose Cartesian displacements $J \Delta \theta_a$, Eq. (5), best fit the observed Δr . We get

$$\Delta \theta = (J^t M J)^{-1} J^t M \Delta r. \quad (6)$$

The matrix $J^t M J = T$ is the kinetic energy Eq. (4). We define the torsional fraction of a Cartesian displacement vector as $\tau(\Delta r) = (J \Delta \theta, M J \Delta \theta) / (\Delta r, M \Delta r)$ where brackets denote scalar product. For CA interactions, the lowest frequency ANM modes are almost completely torsional ($\tau \approx 1$) and very similar to TNM modes. However, as frequency increases, ANM modes become less torsional and more different from TNM modes (see supplementary material, Fig. S1 [26]). For other types of interaction ANM and TNM modes are rather different.

We examined 21 pairs of proteins with large conformation change and identical sequence taken from a large-scale study [27]. Structures were superimposed with the McLachlan algorithm [28] minimizing the mass-weighted deviation $\sum_i m_i |\vec{r}_i^A - \vec{r}_i^B|^2$, which imposes the Eckart con-

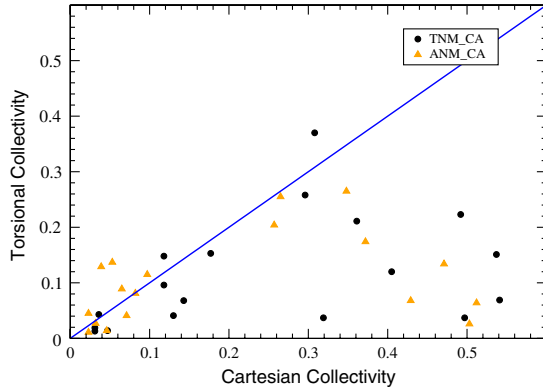


FIG. 1 (color online). Torsional versus Cartesian collectivity of the four lowest frequency normal modes for 21 proteins performing large conformation changes, using TNM and ANM with C_α interactions. The diagonal $y = x$ is plotted for comparison.

ditions. Deviations $\Delta \tilde{r}_i^{AB} \equiv \tilde{r}_i^A - \tilde{r}_i^B$ were projected onto the normal modes of the unbound structure, obtaining squared projections $(c_\alpha)^2 = (\Delta r^{AB}, M \chi^\alpha)^2 / (\Delta r^{AB}, M \Delta r^{AB})$. There is almost always significant correlation between the contribution of each mode α to thermal fluctuations (ω_α^{-2}) and to conformation changes (c_α^2), e.g., $r \in [0.22, 0.99]$ in supplementary Table IX [26], which quantifies the common notion that functional motions of proteins are correlated with their equilibrium dynamics. This is consistent with modeling the conformation change as the linear response of the protein to the perturbation induced by the ligand [29], without calling evolution into play. To get deeper insight, we consider the free energy barrier of the conformation change, which can be roughly estimated as [30]

$$\Delta E^{AB} = \frac{1}{2} \sum_\alpha (c_\alpha \omega_\alpha)^2 (\Delta r^{AB}, M \Delta r^{AB}). \quad (7)$$

ΔE depends on the ratio between the contribution of mode α to the conformation change and to thermal fluctuations. A simple neutral model is $c_\alpha^2 \propto \omega_\alpha^{-2}$, which is the distribution of coefficients with maximum Renyi entropy for given ΔE . We define the excess correlation coefficient $\rho = r((c_\alpha \omega_\alpha)^2, \omega_\alpha^{-2})$ that measures deviations from this neutral model, which predicts $\rho \approx 0$, so that $\rho > 0$ tends to reduce the predicted free energy barrier and $\rho < 0$ tends to increase it. We measured ρ for our 21 proteins pairs, finding that it is significantly positive in 9 cases and nonsignificant in 12. We then measured the reciprocal collectivity of the squared coefficients c_α , $\kappa[(c_\alpha)^2]$, which quantifies the number of normal modes that contribute to the conformation change. In Fig. 2 top we plot this quantity divided by the number of residues versus the excess correlation coefficient ρ . There is an approximately exponential relationship between these quantities; i.e., when ρ is large few low frequency normal modes are sufficient to describe the conformation change. We propose that these two features (low frequency modes contributing more than $1/\omega_\alpha^2$ and very few modes contributing to the conformation change)

might be a signature of selective pressure to reduce the free energy barrier. More work will be needed to test this hypothesis. Nevertheless, consistently with it, we observed that conformation changes involved in transport or enzymatic reactions have larger ρ and are described by fewer normal modes than those that are likely not functional, i.e., a pair of structures determined at different pH (2cg6A-2cg7A), a pair of proteins crystallized in different forms (2g4oC-1w0dD), and a mutant protein that undergoes domain swapping and dimerization (1y51C-1y50A). The pair 1i0cA-1i07B undergoes domain swapping with putative regulatory function. The two pairs involved in domain swapping show the most negative ρ and, consequently, some of the largest barriers and number of relevant modes.

One can see from Fig. 2 that on the average a smaller number of TNM than ANM modes is sufficient to describe the conformation change, in particular, when high frequency modes are important, and that ρ is equal or larger for the TNM than for the ANM. The interaction scheme yielding the lowest reciprocal collectivity for the ANM is the one based on C_α atoms with threshold distance 9.0 Å (mean $\kappa/N_{\text{res}} = 0.291$, mean $\rho = 0.104$), and for the TNM it is the one based on C_β atoms with threshold distance 9.0 Å and backbone atoms as reference (mean $\kappa/N_{\text{res}} = 0.182$, mean $\rho = 0.192$). In all cases, the mean ρ excludes the nonfunctional movements.

Finally, we compare predicted thermal fluctuations with observed conformation changes. The vectors $q = T^{1/2}\theta$ naturally arise in NMA, where $u^\alpha = T^{1/2}v^\alpha$ is the eigen-system of the rescaled potential energy $\tilde{U} = T^{-1/2}U^\theta T^{1/2}$. Predicted thermal fluctuations $\langle q_a^2 \rangle$ significantly correlate with observed $\Delta q_a^2 = (T^{1/2}\Delta\theta)_a^2$ for protein pairs with large ρ , but not for pairs with small ρ . They are shown in color code in Fig. 2 bottom for the enzyme adenylate kinase. Their correlation coefficient is 0.74 for both the TNM and the ANM. Notice that the largest $\langle q_a^2 \rangle$ components are close to the binding site.

In summary, we have introduced the TNM, whose degrees of freedom are torsion angles of the protein backbone. TNM normal modes allow us to accurately reconstruct the positions of more atoms than similar Cartesian ENMs and to define different interresidues interactions, they have a smaller computational cost, and they do not distort the covalent geometry of proteins within first order. Normal modes localized in angular space and collective in Cartesian space may represent hinges in protein motions. TNM modes describe in a more accurate and economic way than the corresponding ANM modes the conformation changes of a protein, in the sense that a smaller number of modes significantly contribute to these motions. Thermal fluctuations predicted through the TNM correlate more strongly with the experimental conformation changes and with experimentally observed B factors, although the latter comparison is complicated by the contribution of rigid body rotations [32]. These results make the TNM a competitive method for analyzing conforma-

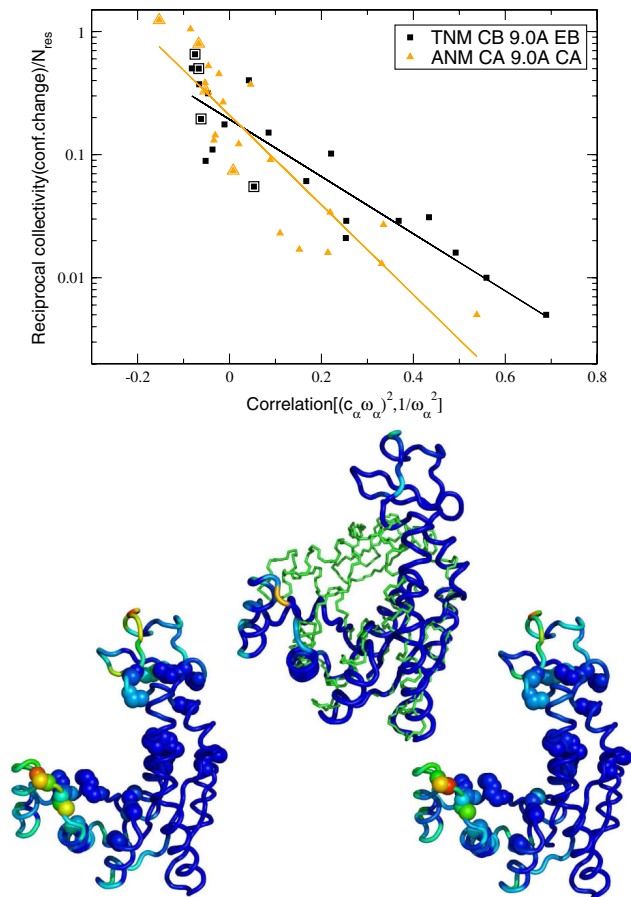


FIG. 2 (color online). Top: For each protein pair, we plot $\kappa[(c^\alpha)^2]/N_{\text{res}}$, estimating the effective number of modes contributing to the conformation change divided by N_{res} , versus $\rho = r[(c_\alpha \omega_\alpha)^2, \omega_\alpha^{-2}]$. The larger ρ , the smaller the predicted energy barrier. Conformation changes that are likely not functional or are regulatory are indicated with concentric symbols. Bottom: Comparison of predicted thermal fluctuations and observed angular conformation change of adenylate kinase (see text). Central figure: open and close conformation (PDB: 4ake, 1ank). Δq_a^2 of the conformation change is shown in color code, from dark (more rigid) to light (more dynamic). Bottom: thermal fluctuations $\langle q_a^2 \rangle$ predicted with the ANM (left) and TNM (right). Functional residues are drawn as spheres [31].

tion changes, with possible applications to flexible docking for drug design and refinement of homology modeled protein structures.

We acknowledge interesting discussions with J. Echave, A. Pascual-García, M. Orozco, and P. Chacón, and funding from the Spanish Science Ministry Grants No. BIO2005-05786, No. BIO2008-04384, No. CSD2006-00023, RyC program (U.B.), and No. SBIO-0214-2006 from Comunidad de Madrid.

*ubastolla@cbm.uam.es

[1] A. del Sol, C.J. Tsai, B. Ma, and R. Nussinov, *Structure* **17**, 1042 (2009).

- [2] N. Tokuriki and D.S. Tawfik, *Science* **324**, 203 (2009).
- [3] C. B-Rao, J. Subramanian, and S.D. Sharma, *Drug Discov. Today* **14**, 394 (2009).
- [4] R. Han, A. Leo-Macias, D. Zerbino, U. Bastolla, B. Contreras-Moreira, and A.R. Ortiz, *Proteins* **71**, 175 (2008).
- [5] R. Goldstein *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).
- [6] M.M. Tirion, *Phys. Rev. Lett.* **77**, 1905 (1996).
- [7] K. Hinsen, *Proteins* **33**, 417 (1998).
- [8] C. Micheletti, P. Carloni, and A. Maritan, *Proteins* **55**, 635 (2004).
- [9] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, and I. Bahar, *Biophys. J.* **80**, 505 (2001).
- [10] I. Bahar and A.J. Rader, *Curr. Opin. Struct. Biol.* **15**, 586 (2005).
- [11] C. Clementi, *Curr. Opin. Struct. Biol.* **18**, 10 (2008).
- [12] F. Tama and Y.H. Sanejouand, *Protein Eng.* **14**, 1 (2001).
- [13] D. Tobi and I. Bahar, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18908 (2005).
- [14] W. Zheng, B.R. Brooks, and D. Thirumalai, *Curr. Protein Pept. Sci.* **10**, 128 (2009).
- [15] S.E. Dobbins, V.I. Lesk, and M.J. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10390 (2008).
- [16] A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, and A.R. Ortiz, *Biophys. J.* **88**, 1291 (2005).
- [17] N. Go, T. Noguti, and T. Nishikawa, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3696 (1983).
- [18] M. Lu, B. Poon, and J. Ma, *J. Chem. Theory Comput.* **2**, 464 (2006).
- [19] J.D. Bryngelson and P.G. Wolynes *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987).
- [20] C. Eckart, *Phys. Rev.* **47**, 552 (1935).
- [21] The vectors \vec{t}_a and \vec{A}_a are obtained through the Eckart conditions $\sum_i m_i \vec{J}_{ia} = 0$ and $\sum_i m_i \vec{r}_i^0 \times \vec{J}_{ia} = 0$. We get $I \vec{A}_a = -I_a \vec{e}_g + M_a (\vec{R} - \vec{R}_a) \times (\vec{e}_a \times (\vec{R}_a - \vec{s}_a))$ and $\vec{t}_a = -\vec{e}_a \times \vec{R} - (M_a/M) \vec{e}_a \times (\vec{R}_a - \vec{s}_a)$, where $M = \sum_i m_i$, $M_a = \sum_{i \in \Omega_a} m_i$, $M_a \vec{R}_a = \sum_{i \in \Omega_a} m_i \vec{r}_i^0$, I is the inertia tensor and I_a is its restriction to the set Ω_a .
- [22] Q. Yang and K.A. Sharp, *Proteins* **74**, 682 (2009).
- [23] S. Omori, S. Fuchigami, M. Ikeguchi, and A. Kidera, *J. Comput. Chem.* **30**, 2602 (2009).
- [24] R. Bruschweiler, *J. Chem. Phys.* **102**, 3396 (1995).
- [25] M. Gerstein and W. Krebs, *Nucleic Acids Res.* **26**, 4280 (1998); R. Mosca, B. Brannetti, and T.R. Schneider, *BMC Bioinf.* **9**, 352 (2008).
- [26] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.104.228103> for the list of conformation changes, detailed results, and plots of the torsional fraction of ANM modes.
- [27] W. Nishima, G. Qi, S. Hayward, and A. Kitao, *Bioinformatics* **25**, 628 (2009).
- [28] A.D. MacLachlan, *Acta Cryst. A* **38**, 871 (1982).
- [29] M. Ikeguchi, J. Ueno, M. Sato, and A. Kidera, *Phys. Rev. Lett.* **94**, 078102 (2005); J. Echave, *Chem. Phys. Lett.* **457**, 413 (2008); J. Echave and F.M. Fernández, *Proteins* **78**, 173 (2010).
- [30] O. Miyashita, P.G. Wolynes, and J.N. Onuchic, *J. Phys. Chem. B* **109**, 1959 (2005).
- [31] Images produced with Pymol <http://www.pymol.org/>.
- [32] M.F. Thorpe, *Phys. Biol.* **4**, 60 (2007); M. Rueda, P. Chacón, and M. Orozco, *Structure* **15**, 565 (2007).